

ERGA Assembly Report

v24.10.15

Tags: ATLASEa[INVALID TAG]

TxID	1838238
ToLID	xgLeuNass1
Species	Leucozonia nassa
Class	Gastropoda
Order	Neogastropoda

Genome Traits	Expected	Observed
Haploid size (bp)	3,606,911,098	3,990,886,037
Haploid Number	31 (source: ancestor)	35
Ploidy	2 (source: ancestor)	2
Sample Sex	Unknown	Unknown

EBP metrics summary and curation notes

Obtained EBP quality metric for collapsed: 6.8.Q44

The following metrics were automatically flagged as below EBP recommended standards or different from expected:

- . Observed Haploid Number is different from Expected
- . Kmer completeness value is less than 90 for collapsed
- . BUSCO single copy value is less than 90% for collapsed
- . BUSCO duplicated value is more than 5% for collapsed

Curator notes

. Interventions/Gb: 75
. Contamination notes: ""
. Other observations: "The assembly of *Leucozonia nassa* (xgLeuNass1) is based on 27X ONT data and 148X Arima Hi-C data generated as part of the ATLASEa programme (<https://www.atlasea.fr>). The assembly process included the following steps: initial ONT assembly generation with Hifiiasm, removal of contaminant sequences using Context, removal of haplotypic duplications using purge_dups, and Hi-C-based scaffolding with YaHS. In total, 29 contigs were identified as contaminants (bacterial, archaeal, or viral), totaling 1.08 Mb (with the largest being 0.139 Mb). Additionally, 3696 regions totaling 343.018 Mb (with the largest being 0.677 Mb) were identified as haplotypic duplications and removed. Mitochondrial genome was assembled using ptGAUL. Finally, the primary assembly was analyzed and manually improved using Pretext. During manual curation, 61 haplotypic regions were removed, totaling 25.134Mb (with the largest being 1.907Mb). Chromosome-scale scaffolds confirmed by Hi-C data were named in order of size. "

Quality metrics table

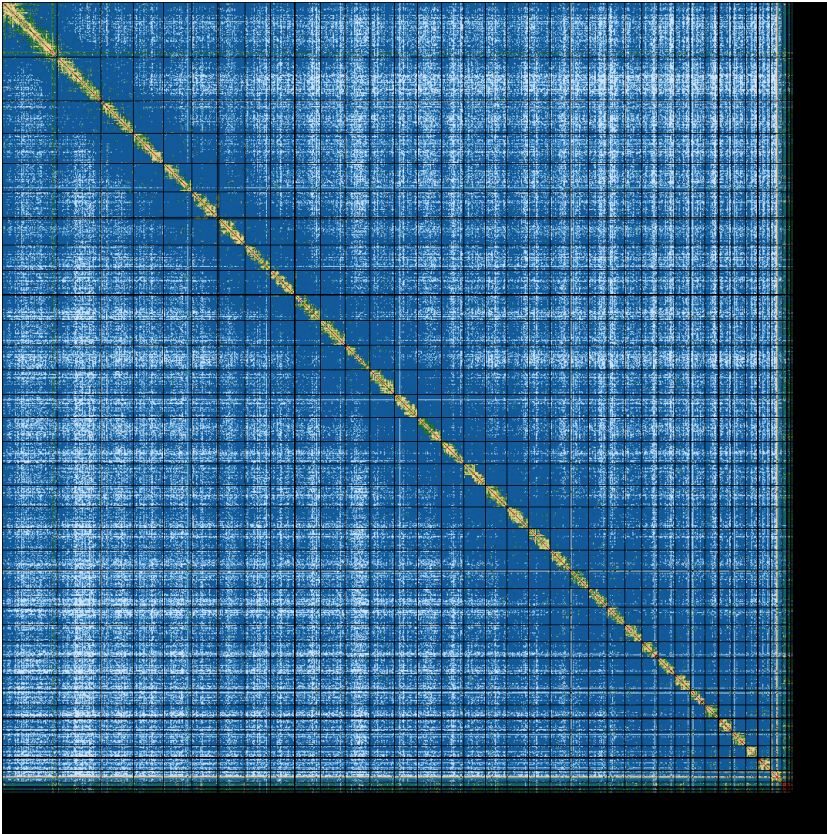
Metrics	Pre-curation collapsed	Curated collapsed
Total bp	4,016,048,457	3,990,886,037
GC %	43.51	43.51
Gaps/Gbp	793.07	802.58
Total gap bp	318,500	333,400
Scaffolds	1,187	1,148
Scaffold N50	109,651,362	111,222,188
Scaffold L50	14	15
Scaffold L90	33	33
Contigs	4,372	4,351
Contig N50	2,150,951	2,150,951
Contig L50	542	538
Contig L90	1,909	1,899
QV	44.0265	44.0228
Kmer compl.	82.7268	82.5422
BUSCO sing.	79.7%	78.1%
BUSCO dupl.	17.4%	16.1%
BUSCO frag.	1.1%	3.5%
BUSCO miss.	1.9%	2.3%

Warning! BUSCO versions or lineage datasets are not the same across results:

BUSCO: 6.0.0 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

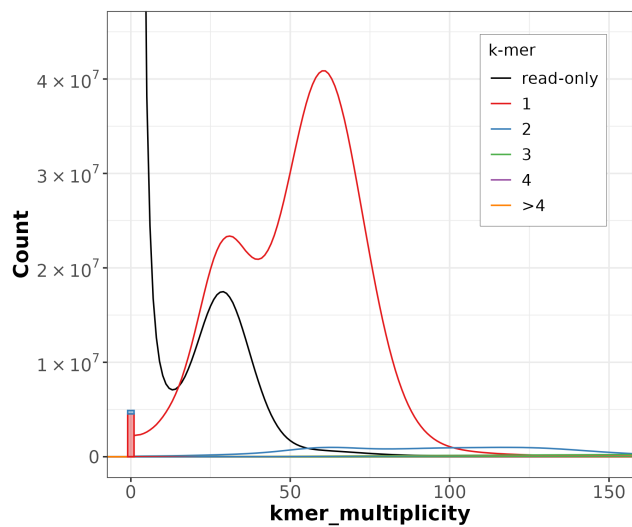
BUSCO: 5.8.2 (euk_genome_min, miniprot) / Lineage: mollusca_odb12 (genomes:36, BUSCOs:4421)

HiC contact map of curated assembly

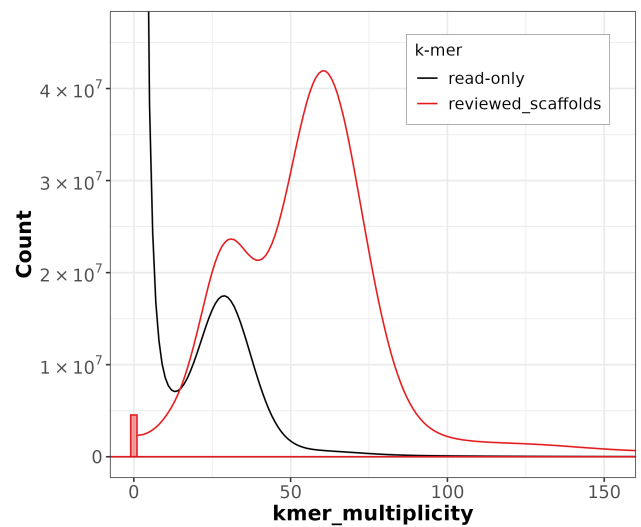


collapsed [\[LINK\]](#)

K-mer spectra of curated assembly

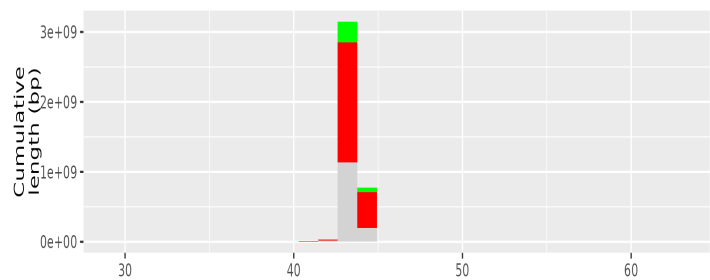


Distribution of k-mer counts per copy numbers found in asm



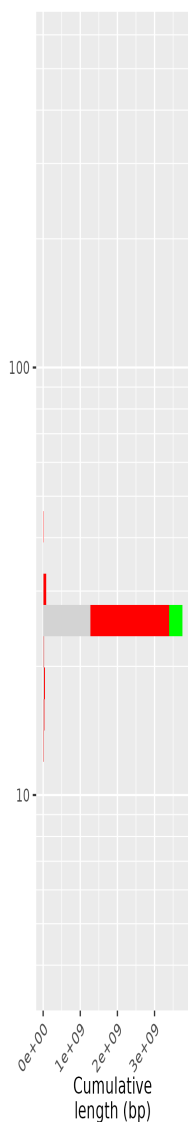
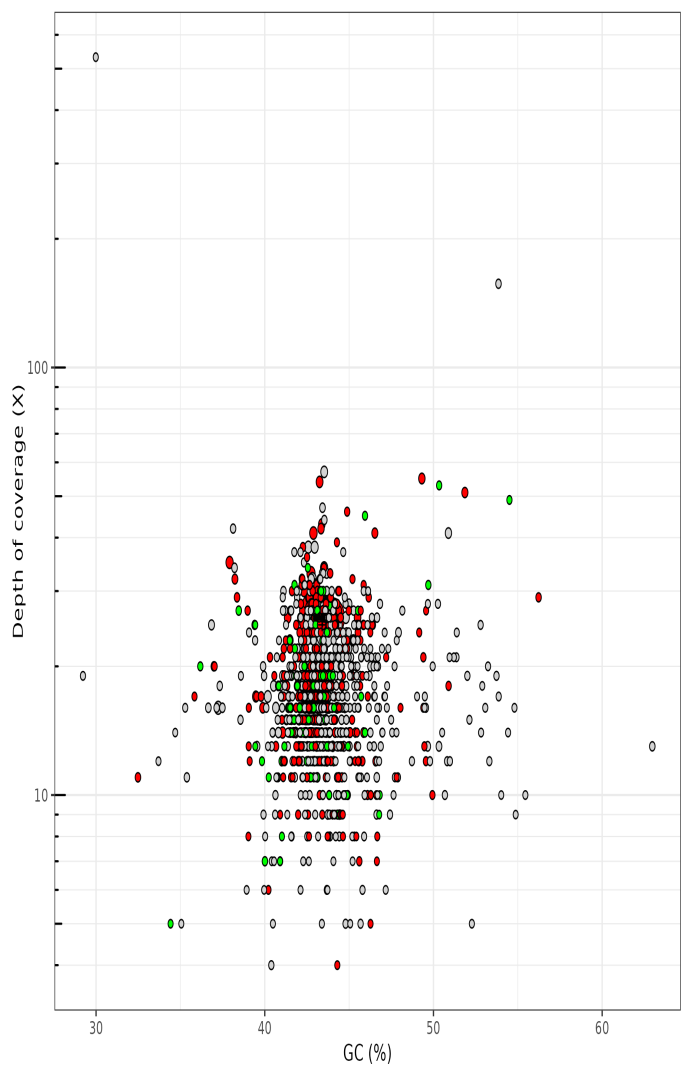
Distribution of k-mer counts coloured by their presence in reads/assemblies

Post-curation contamination screening



TAPAs summary Graph

(3 0X contigs have been hidden)



superkingdom

- Bacteria
- Eukaryota
- N/A

Longest sequences (bp)

- xgLeuNass1_1 - 263209168 (Eukaryota)
- ▲ xgLeuNass1_2 - 212859287 (Eukaryota)
- xgLeuNass1_3 - 152173215 (Eukaryota)
- + xgLeuNass1_4 - 147259752 (Eukaryota)
- ▣ xgLeuNass1_5 - 136948366 (Eukaryota)

Length (bp)

- 5.0e+07
- 1.0e+08
- 1.5e+08
- 2.0e+08
- 2.5e+08

collapsed. Bubble plot circles are scaled by sequence length, positioned by coverage and GC proportion, and coloured by taxonomy. Histograms show total assembly length distribution on each axis.

Data profile

Data	Long reads	Arima
Coverage	27	149

Assembly pipeline

- **Hifiasm**
 - |_ *ver*: 0.19.5-r593
 - |_ *key param*: NA
- **purge_dups**
 - |_ *ver*: 1.2.5
 - |_ *key param*: NA
- **YaHS**
 - |_ *ver*: 1.2
 - |_ *key param*: NA

Curation pipeline

- **PretextMap**
 - |_ *ver*: 0.1.9
 - |_ *key param*: NA
- **PretextView**
 - |_ *ver*: 0.2.5
 - |_ *key param*: NA

Submitter: Arnaud Couloux

Affiliation: Genoscope

Date and time: 2026-08-27 12:10:56 CEST